**Mobile Data Collector Based Data Gathering Technique in a Cluster Based Sensor Network**

*A thesis submitted in partial fulfilment of the requirement*

*for the award of degree of*

**BACHELOR OF TECHNOLOGY**

**IN**

**INFORMATION TECHNOLOGY**

*Submitted by*

**ABHILASH MAJUMDER**

**(15/IT/02)**

*Under the supervision of*

**DR. SANGHITA BHATTACHARJEE**

****

**Department of Computer Science and Engineering**

**National Institute of Technology, Durgapur**

**West Bengal, India**

**December, 2018**

****

**CERTIFICATE**

This is to certify that the work contained in the thesis entitled **“Mobile Data Collector Based Data Gathering Technique in a Grid Based Sensor Network”** by **Abhilash Majumder** has been carried out under the guidance of **Dr. Sanghita Bhattacharjee**. The results reported herein are original and have not been submitted to any other University or Institute for the award of any degree or diploma.

**Dr. Goutam Sanyal Abhilash Majumder(15/IT/02)**

Professor and Head of the Dept.,

CSE Department,

NIT Durgapur

**Examiners Dr. Sanghita Bhattacharjee**

Assistant Professor,

1. CSE Department,

NIT Durgapur

2.

3.

**ACKNOWLEDGEMENT**

I ,the undersigned, would like to share this space in thanking **Dr. Sanghita Bhattacharjee** ,assistant professor of department of Computer Science and Engineering and **Dr.Goutam Sanyal,** professor and head of department of Computer Science and Engineering NIT Durgapur for the wonderful opportunity provided to me to pursue research in the domain of wireless sensor networks. I would also like to share this space to thank family and friends for their constant support.

**Abhilash Majumder**

**(15/IT/02)**

Department of Information Technology

NIT Durgapur

December 09 2018

**ABSTRACT**

In Wireless Sensor Networks (WSN), energy consumption of the nodes is a very critical issue since sensor nodes are battery operated and it is very difficult to replace the batteries when they are deployed in harsh conditions. In this project, a mobile data collector is used for gathering data to prolong the lifetime of the sensor network. The field is divided into cluster by k means algorithm which contain sensors .The k depends on the user and more the value of k changes the outcome of energy and data collection rate of the total sensor cluster. Each cluster contains a Euler circuit of the sensors which are then used for average data collection rate and energy, lifetime calculation. Intra and inter cluster analysis has also been done with added latency in transmission time.

**TABLE OF CONTENTS**

**Certificate 2**

**Acknowledgement 3**

**Abstract 4**

**Table of Contents 5**

**List of Figures 6**

**List of Tables 7**

**Chapter 1: Introduction**

1.1 Wireless Sensor Networks

1.2 Clustering Approaches

1.3 Grid-Based Approaches

1.4 K-means Approach

1.5 Advantage of K –means over Grid system

1.4 Contribution of the Work

**Chapter 2: Related Work**

**Chapter 3: Proposed Work**

3.1 Introduction and Assumptions

3.2 Important Terms

3.3 Algorithmic Description

**Chapter 4: Graphical Analysis**

**Chapter 5: Conclusion and Bibliography**

**OBJECTIVE**

The objective of the project is to generate maximum data flow for the longest interval in a cluster of sensors and optimizing the entire system to reduce latency and find better path algorithms. The project uses a rigorous k –means approach to calculate the nearest members to form clusters. After clusters are formed they are given latencies of travel time by collector head. Intra cluster data information rate, energy consumption are calculated by applying Euler tour of the path which is dynamically updated with increase in time. For inter cluster data rate the transmitted data is collected in head/collector while in that time being another clusters gather data. The major topic of interest is which algorithm simulates the entire event before the sensors get thrashed with latency and overflow data. The objective of the project lies in identifying the optimal positioning of the sensors and clustering them in appropriate clusters such that a closed loop exists and the sensors can provide information for long time without falling prey to deadlocks and starvation in the network.

**CHAPTER 1: INTRODUCTION**

**1.1 Wireless Sensor Networks**

Wireless sensor network (WSN) refers to a group of spatially dispersed and dedicated sensors for monitoring and recording the physical conditions of the environment and organizing the collected data at a central location. WSNs measure environmental conditions like temperature, sound, pollution levels, humidity, wind, and so on. Sensor nodes sense the area and send sensed data to base station via single or multi hop communication. Due to the small size and low cost of the sensor nodes, they have a very rapid and large application area like military surveillance, environment monitoring, agriculture, health monitoring, etc.

Sensor nodes once deployed in field work unattended. They are highly resource constrained as they have limited processing speed, storage capacity and energy in their on-board battery. All processing done by sensor nodes (sensing, data processing and data communication) is energy consuming. The scarce energy resource and the harsh environmental conditions make replacement or recharging of the battery practically impossible in certain situations like battle field, volcano detection and deep-sea sensing. Thus, energy efficiency of nodes is a key design issue for wireless sensor networks.

* + 1. **Characteristics of WSN**
* Sensor nodes are densely deployed. So multiple sensors are deployed to measure the same or similar physical phenomenon.
* Sensor nodes are prone to failure because of battery exhaustion and harsh environmental conditions.
* Node failure often changes the topology of a sensor network.
* Communication failures are also a critical issue in WSN.
* Due to small size, sensor nodes are limited in power, computational capabilities and memory.
  + 1. **Applications of WSN**
* Health care monitoring: The sensor networks for medical applications can be of several types: implanted, wearable, and environment-embedded. Possible applications include body position measurement, location of persons, overall monitoring of ill patients in hospitals and at home.
* Area monitoring: Area monitoring is a common application of WSNs. In area monitoring, the WSN is deployed over a region where some phenomenon is to be monitored. A military example is the use of sensors detect enemy intrusion; a civilian example is the geo-fencing of gas or oil pipelines.
* Air pollution monitoring: Wireless sensor networks have been deployed in several cities to monitor the concentration of dangerous gases for citizens These can take advantage of the ad hoc wireless links rather than wired installations, which also make them more mobile for testing readings in different areas.
* Forest fire detection: A network of Sensor Nodes can be installed in a forest to detect when a fire has started. The nodes can be equipped with sensors to measure temperature, humidity and gases which are produced by fire in the trees or vegetation. The early detection is crucial for a successful action of the firefighters; thanks to Wireless Sensor Networks, the fire brigade will be able to know when a fire is started and how it is spreading.
* Landslide detection: A landslide detection system makes use of a wireless sensor network to detect the slight movements of soil and changes in various parameters that may occur before or during a landslide. Through the data gathered it may be possible to know the impending occurrence of landslides long before it actually happens.
* Water quality monitoring: Water quality monitoring involves analyzing water properties in dams, rivers, lakes and oceans, as well as underground water reserves. The use of many wireless distributed sensors enables the creation of a more accurate map of the water status and allows the permanent deployment of monitoring stations in locations of difficult access, without the need of manual data retrieval.
* Machine health monitoring: Wireless sensor networks have been developed for machinery condition-based maintenance (CBM) as they offer significant cost savings and enable new functionality.Wireless sensors can be placed in locations difficult or impossible to reach with a wired system, such as rotating machinery and untethered vehicles.
  1. **Clustering Approaches**

A cluster-based sensor network consists of related sensor nodes grouped into clusters. There is a cluster head (CH) for each cluster. It manages operation of the nodes in a cluster. The individual nodes within a cluster send their sensed data to their corresponding cluster head. The cluster heads aggregate the received data, perform data compression if necessary and then send the data to the base station directly or through intermediate cluster heads resulting in multihop transmission.

An important issue in cluster-based sensor networks is the *hotspot* problem, in which CHs close to the base station have a huge relay load as they collect data not only from its own cluster members but also from other CHs and forward them to the base station. This results in rapid energy depletion of such nodes causing network partitioning. As a result, even though nodes far away from the sink may have sufficient energy left, their energy would not be utilized as the nodes near the sink cannot relay their data to the sink because their energy has been depleted. This significantlyreduces the lifetime of the network.

* 1. **Grid-Based Approaches**

In grid-based approaches, the entire area of sensing is divided into virtual grids and thus different sensor nodes may belong to different grids. The grids may be equal sized square shaped grids as it has been proposed in the current project or may be of variable sizes. The nodes in each grid are determined by their coordinates. Various strategies exist for data transmission in such networks. A common approach is to select a cluster head (CH) for each grid which transmit the data to the base station as in the case of traditional clustering approaches. Grid-based techniques are popular due to their simplicity, scalability and uniformity in energy consumption across the network.

One important issue in grid-based sensor networks is determining the number of grids and suitable grid size. Often, it becomes difficult to achieve the desired number of grids for a particular deployment scenario. Moreover, the network performance is affected in case of nonuniform deployment.

* 1. **K-means Approach**

Grid system often used to determine the number of sensor networks, falls prey to common disadvantage that the system fails to generate optimal grid layout.This is solved by using a machine learning approach to generating clusters for sensor networks-means is such an algorithm which considers nearest neighborhoods of clusters within a designated specified range.The supervised algorithm takes into consideration nearest distance vectors of the individual sensors and sorts them in accordance of their latency time and residual energies. The greater the similarity in the overall energy demand of two clusters the greater the match .The K –means uses the data information rate to predict the next cluster inside the system. Algorithms like Euler and Hamiltonian circuit is used to determine the nearest neighbor.

* 1. **Advantage of K-means over Grid System**

The advantage of this algorithm is that it uses machine learning to simulate the events and clusters. This algorithm uses previous k means value and predicts the nearest value of data information with the new sensors which are again grouped into another cluster. This algorithm is slow for small sets but runs with very high optimality and complexity for large number of clusters.

* 1. **Contribution of the Work**

In this project, a mobile data collector-based data gathering technique has been proposed in a grid-based wireless sensor network. The aim is to maximize the network lifetime. The field is initially divided into equal sized square shaped grids. The member nodes of each grid are determined by their coordinates. The grids are then merged with their neighboring grids based on the average energy density of the grids, *i.e.*, the ratio of the sum of residual energies of all the member nodes and the number of member nodes. This results in formation of super-grids consisting of multiple smaller grids. The centroids of these merged grids are then used as the points for data collection by the mobile data collector. A suitable path for data collection covering all these merged grids is determined by taking into account the average residual energies of the merged grids and the distances between their centroids. Extensive simulations are performed to demonstrate the effectiveness of the proposed algorithms.

**CHAPTER 2: RELATED WORK**

Various clustering as well as grid-based approaches exist which aim to maximize the network lifetime.

In [1], the authors have proposed LEACH, a clustering-based protocol that utilizes randomized rotation of cluster heads to evenly distribute the energy load among the sensors in the network. CH depends on a random number between 0 and 1. If the selected number is less than a threshold (which depends on the percentage of cluster heads), then the node becomes a cluster head for the current round.

In [2], LEACH-C was proposed which is a modified version of LEACH. The number of CHs is determined by base station and varies from round to round due to lack of coordination among the nodes. In LEACH-C, the no of cluster heads in each round equals a predetermined optimum value.

In [3],the authors have proposed a Grid Based Hybrid Network Deployment Approach (GHND) in which the field is divided into equal sized square-shaped grids. It uses merge and split technique to achieve load balancing. Nodes are merged with neighboring grids on the basis of Weighted Merge Score (WMS). Four splitting strategies are discussed to split the zone if the number of nodes exceed the upper bound. After the topology construction, zone head (ZH) is selected on the basis of Average Distance Value (ADV). The ZHs are rotated to increase the network stability and lifetime.

In [4], the authors have proposed a Mobile Sink based Routing Protocol (MSRP) for prolonging the lifetime in a clustered WSN. It aims to overcome the hotspot problem by using a mobile sink is used to sense data from the CHs in its vicinity. The sink keeps track of the residual energies of all CHs and moves to the CHs having higher energy. Hence, the hotspot problem is minimized as due to sink movement, the responsibility of relaying the data is shared by different high energy CHs near the sink.

In [5], a similar mobile sink-based approach called Energy Efficient Mobile Sink Routing Algorithm (EEMSRA) has been proposed in which the sink moves based on the average energy in each cluster. The clusters are formed using the LEACH algorithm.

In [6], the authors have proposed a mobile sink based approach for collecting sensed data from CHs. The optimal visiting points and the data gathering path is determined. It uses k-means clustering algorithm to form clusters and find the optimal visiting points. To minimize the data gathering time, an optimal travel path is selected for the mobile sink before every data gathering tour using some solutions of MST (Minimum Spanning Tree), Euler cycle and Hamiltonian cycle.

**CHAPTER 3: PROPOSED WORK**

**3.1 Introduction and Assumptions**

Instead of having grid division of the entire layout of the sensors ,the K means algorithm applies a more granular effect and decides intrinsically which sensors to include in which cluster.

K-means ([MacQueen, 1967](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html#macqueen)) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.  
Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image009.gif ,

where https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image011.gif is a chosen distance measure between a data point https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image013.gif and the cluster centre https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/images/image015.gif, is an indicator of the distance of the n data points from their respective cluster centres. Since the value k partitions have been given before ,the algorithm runs sub optimally and is generally NP hard.

This work adopts the energy consumption model used in the first order radio model [1]. Energy is consumed within each round of data transmission. The length of each data packet is fixed, say l bits. The energy spent in transmitting a l-bit message a distance d using first order radio model is given by the following equation:

ETx = Eelec+ Eamp\*l\*d\*d (3.1)

Here, the distance d would be the distance between a sensor node and the position of the data collector, *i.e.*, the centroid of the current cluster.

During data gathering, the mobile data collector has to wait at the collection point of a particular grid for a sufficient time so that data is collected from all the nodes belonging to that grid. The waiting time in a grid for data packet of size l and channel capacity R is given by:

Twaiting= (l/R) \* number of nodes in the grid (3.2)

It is also assumed that there is no energy loss due to collision of packets and that the mobile data collector has sufficient memory to store the data packets received from all the sensor nodes in a round.

L

L

**Figure 3.1: Pictorial representation of the cluster formation (k=4)**

It is also assumed that energy is spent only in transmission of data and no energy is consumed during sensing or reception of data by the mobile data collector.

**3.2 Important Terms**

* **Network Lifetime:**It is defined as the number of rounds after which the first node dies.
* **Latency:** It is the average time taken by the mobile data collector to complete each round of data collection.

**3.3 Algorithmic Description**

The proposed technique is divided into five phases, namely cluster division by k-means, validity of Euler Circuit in the graph of clusters, energy consumption and data information attained from single sensor, data collection from multiple sensors in a cluster, inter cluster transmission of data and collector time latency.

**3.3.1 K –Means Algorithm**

The k means uses initial information of the data packets and also the energy values.The algorithm starts with placing k group points or initial centroids in the entire sensor area .Each sensor is added to that cluster which has the nearest centroid.When all the sensors have been divided into clusters recalculate the position of the k centroids and the process is repeated.

**Algorithm 1: K-means algorithm**

**Input:** List of sensor nodes *node*, initial value of k

**Begin**

Initialise k means between maximum gathering of sensors and minimum gathering within a region of distance x

**For** a given number of iterations

Find Euclidean mean distance between each sensor to respective values of centroids

Assign the sensor to that cluster

Update the mean/average

**End Loop**

**End**

**3.3.2 Validity of Euler Circuit in Graph**

After clusters have been decided the Euler path can be checked by using traditional Graph algorithms. An undirected graph has Eulerian cycle if following two conditions are true.

**Eulerian Cycle**

a) All vertices with non-zero degree are connected. We don’t care about vertices with zero degree because they don’t belong to Eulerian Cycle or Path (we only consider all edges).  
b) All vertices have even degree.

**Eulerian Path**  
An undirected graph has Eulerian Path if following two conditions are true.  
a) Same as condition (a) for Eulerian Cycle

b) If two vertices have odd degree and all other vertices have even degree. Note that only one vertex with odd degree is not possible in an undirected graph (sum of all degrees is always even in an undirected graph)

Note that a graph with no edges is considered Eulerian because there are no edges to traverse.  
In Eulerian path, each time we visit a vertex v, we walk through two unvisited edges with one end point as v. Therefore, all middle vertices in Eulerian Path must have even degree. For Eulerian Cycle, any vertex can be middle vertex, therefore all vertices must have even degree

**Algorithm 2: Validity of Euler Path**

**Input:** Initialise the clusters and sensors as nodes in a graph model with undirected edges between them.

**Begin**

Form the graph by adding edges

Check for connected component in graph

Find a vertex with non zero degree

**If** no such edge return true

**Else** do DFS of the graph

Check if all non zero degree nodes are connected

Count the vertices with odd degree

**If** count is more than two graph is non eulerian

**Else if**  count is two graph is semi Euler

**Else if** count is 0 graph is Euler

**End**

**3.3.3 Energy Consumption from Single Sensor**

The energy consumption from a sensor has been calculated as product of the packet information rate with the B, where B is (L/R) ,L is length of channel and R is the channel capacity.After a Euler tour of the circuit is obtained each sensor has an initial energy value .The energy consumed during its gathering of data is

(Eelec+ Eamp\* l \* d \* d)\*x\*B, where x is the packet information rate .Residual energies are also calculated .Lifetime of each sensor is calculated by dividing the residual energy by total energy consumed.

**Algorithm 3: Energy Consumption from single sensor**

**Input:** Graph containing a valid Euler tour circuit with necessary k means cluster

**Begin**

Initialise Eamp and Eel values, channel capacity, channel length and packet rate.

Check for Euler Tour in the Graph

Store the order of tour in a container of pairs of start node and end node

Calculate the packet information rate

**For** each edge

Attain the energy consumption

Determine the residual energy

**If** energy of the sensor becomes negative remove it from graph

**Else**

Calculate the lifetime of the sensor

**End loop**  when all the sensors are exhausted or the all edges are calculated

**End**

**3.3.4 Energy and Data Collection from multiple sensors in a cluster**

Energy and data collection from multiple sensors in a cluster(intra cluster analysis) is done dynamically. The reason is that while a sensor donates its data to the collector head other sensors also carry on gathering information ,they donot remain idle. The information packet rate also changes by a fraction of original value for each cycle .As a result the sensors all thrashed or exhausted after finite number of cycles. Here also optimal Euler path needs to be decided between the sensors in a cluster with minimized latency. In that latency time, which includes data gathering by the head/collector and moving over to another sensor in the cluster, more information are collected by the other waiting sensors.

**Algorithm 4: Intra cluster energy and data collection**

**Input:** Graph contains Euler cycle and also contains values of Eamp , Eel, d , l, channel capacity, packet information rate, channel length

**Begin**

Store the order of Euler tour in a container

Calculate the maximum value (T) of the packet information rate attained

**For** time =0 to T

**Do** calculate the energy of each sensor

Residual energy of the sensor

**If**  the sensor has lost energy remove it from graph

Calculate the lifetime

Calculate the data collected for n sensors in a cluster as

X\*(L/R)\*n where X is packet information rate

Update value of X(packet information rate)

Add latency time and transmission time to T

**End Loop**  when either all the sensors exhaust or packet information rate is highest

**End**

**3.3.5 Inter Cluster Analysis**

Inter cluster analysis is primarily done between two clusters containing different number of sensors with the optimisation that the collector head gathers information from all the sensors without causing starvation of other sensors.Inter and intra Euler path have been constructed for the algorithm to work and latency time for inter cluster transmission of head/collector is also taken into consideration.Energy and data collection analysis has been done .

**Algorithm 5: Intra Cluster Energy and Data Collection**

**Input:** Euler Graph of k means(k=2) sensors have been constructed and checked for valid path. Channel capacities and bandwidth and sensor position of each cluster has been recorded.

**Begin Loop** until all the sensors in single/both cluster exhaust

For first cluster

Store the order of Euler tour in a container

Calculate the maximum value (T) of the packet information rate attained

**For** time =0 to T

**Do** calculate the energy of each sensor

Residual energy of the sensor

**If**  the sensor has lost energy remove it from graph

Calculate the lifetime

Calculate the data collected for n sensors in a cluster as

X\*(L/R)\*n where X is packet information rate

Update value of X(packet information rate)

Add latency time and transmission time to T

Average data collected in n point sensors X\*X\*(L/R)\*n

**End Loop**  when either all the sensors exhaust or packet information rate is highest

T1= time to complete first cluster

For second cluster

Increase time to T2 where T2= data collection time from first cluster + latency transmission time

Calculate the packet information rate attained,and average value of data collected .

**For** time =T1 to T2

**Do** calculate the energy of each sensor

Residual energy of the sensor

**If**  the sensor has lost energy remove it from graph

Calculate the lifetime

Calculate the data collected for m sensors in a cluster as

X\*(L/R)\*m where X is packet information rate

Update value of X(packet information rate)

Add latency time and transmission time to T

Average data collected in n point sensors X\*X\*(L/R)\*m

**End Loop**  when either all the sensors exhaust or packet information rate is highest

T2= time to complete second cluster

Again go to first cluster

Increase time to T3 where T3= data collection time from second cluster + latency transmission time to go first cluster

Calculate the packet information rate attained,and average value of data collected .

**End**

**CHAPTER 4: SIMULATION RESULTS**

In this chapter, the performance of the proposed algorithm is evaluated through various performance metrics.

**4.1 Simulation Environment**

All simulations are carried out using C++ programming language on Windows 10 Operating System. To analyze the energy consumption of the nodes, first order radio model [1] is used.

The list of parameters used for simulation is given in Table 4.1

Table 4.1: Simulation Parameters

|  |  |
| --- | --- |
| **Parameters** | **Values** |
|  |  |
| Area | 500\*500 (in m2) |
| No. of Nodes | 100-1000 |
| Distribution | Random |
| No. of sinks | 1 |
| No. of mobile data collectors | 1 |
| Length of data packet | 1000 bits |
| Initial Energy of Nodes | 5-10 J |
| Speed of Mobile Data Collector | 20 m/s |
| Packet Generation Rate | 0.02 packets/s |
| Channel Capacity (R) | 1024 kbps |
| Eelec | 50 nJ/bit |
| Eamp | 100 pJ/bit/m2 |
| Position of the sink | Center of the field, *i.e.*, (250, 250) |

**4.2 Results**

**4.2.1 Plot of Lifetime vs Energy**

The plot shows the variation of cluster lifetime with that of Energy or energy factor alpha. Alpha tends to 1 when the value of lifetime is reduced .The energy for large lifetime of clusters are greater and this signifies that sensors containing large energy can continue to gather information for longer duration without exhausting.

Fig 4.1

**4.2.2 Plot of Percentage of Alive Nodes vs No. of Rounds**

Figure 4.7 shows the plot of percentage of alive nodes as the no. of rounds increases. After a certain no. of rounds, given by the lifetime of the network, nodes start to die off as their residual energy becomes 0. It is observed that the percentage of alive nodes steadily decreases as the no. of rounds increases.

Fig. 4.7

**CHAPTER 5: CONCLUSION AND FUTURE WORK**

**5.1 Conclusion**

In this project, algorithms have been proposed for data gathering using a mobile data collector in a cluster-based wireless sensor network to improve the lifetime and latency. The clusters are generated by k means are further optimized by applying k means when a node sensor dies .This leads to greater output of data and longevity as the edge from the graph containing dead nodes are removed ,which changes overall k means by reducing it. However k means may not always give optimal solution as it is an unsupervised model and in NP hard .Computational algorithms like Travelling Salesman Problem and modified distribution method /assignment method algorithms may offer better solutions than k means. Another algorithm called K nearest neighbors can be also be used. The main topic also lies in identifying if Euler tour always gives sufficient path which is always optimized. Since we are considering sensor nodes in the vicinity of each other and clustering them with k means and creating the graph, there also lies an exclusion of the fact that there may not be a constructive graph to work upon as there may be nodes which are sufficiently further from each other to form a graph.

**BIBLIOGRAPHY AND LINK**

[1] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan,“Energy-efficient communication protocol for wireless microsensornetworks,” in *Proceedings of the IEEE 33rdAnnualHawaiiInternational Conference on System Sciences (HICSS ’00)*, vol. 2,p. 10, January 2000.

[2] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan,“An application-specific protocol architecture for wirelessmicrosensor networks,” *IEEETransactions onWireless Communications*,vol. 1, no. 4, pp. 660–670, 2002.

[3] Haleem Farman, Huma Javed, Jamil Ahmad, Bilal Jan, and Muhammad Zeeshan,"Grid-Based Hybrid Network Deployment Approach forEnergy Efficient Wireless Sensor Networks", Published on 28 August 2016.

[4] Nazir, B., &Hasbullah, H. (2010). Mobile sink based routing protocol (MSRP) for prolonging networklifetime in clustered wireless sensor network. In 2010 International conference on computer applicationsand industrial electronics (ICCAIE 2010).

[5] Yuan, X., & Zhang R. (2011). An energy-efficient mobile sink routing algorithm for wireless sensornetworks. In Wireless communications, networking andmobile computing (WiCOM), 2011 7thinternational conference.

[6] Ilkyu Ha,MamurjonDjuraev, ByoungchulAhn, An Optimal Data Gathering Method for MobileSinks in WSNs, Published on 15 June 2017.

**Repository on github for reference:**

https://github.com/abhilash1910/K-means-clustering-for-WSN